

# Praktikum z ekonometrie

VŠE Praha

*Tomáš Formánek*

# Block 4 – Linear mixed effect models – Outline

- 1 Introduction
- 2 LME model
- 3 Complex LME models
- 4 LME models – references, R packages

# Introduction

LME generalization of a linear regression model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R}),$$

where

$\mathbf{y}$  is a vector of dependent variable observations,

$\mathbf{X}$  is a  $(N \times p)$  matrix of “fixed effects” with  $N$  observations and  $p$  predictor variables,

$\boldsymbol{\beta}$  is a vector of “fixed-effects” regression coefficients,

$\mathbf{Z}$  is a  $(N \times s)$  design matrix for the  $s$  “random effects”  $\mathbf{u}$  that are complementary to  $\boldsymbol{\beta}$ ,

$\boldsymbol{\varepsilon}$  is the error term,

$\mathbf{u}$  and  $\boldsymbol{\varepsilon}$  are assumed normally distributed and mutually independent, with variance-covariance matrices  $\mathbf{G}$  and  $\mathbf{R}$  respectively.

- **Fixed effects:** variables with expected impacts (effects) on the dependent/response variable.

Essentially the same as explanatory variables in a standard linear regression models.

- **Random effects:** typically, grouping factors that we are trying to control for.

RE are always associated with categorical variables (R cannot treat continuous variables as random effects).

Often, we are not interested in specific impacts of individual random effects on the response variable – but we know RE might be influencing the patterns we see & we want to control for such processes.

LME generalization of a linear regression model – example:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

If observations  $i = 1, \dots, N$  are organized into  $j = 1, \dots, J$  relevant groups, and assuming there are random effects on both  $\beta$ -coefficients, we may generalize the model into a LME (after re-arranging):

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{ij} + \varepsilon_{ij},$$

where the combined  $ij$  subscript refers to an  $i$ th observation that belongs to a  $j$ th group.

In the LME model,  $u_{0j}$  and  $u_{1j}$  are stochastic deviations from  $\beta$ -coefficients that are associated with a particular  $j$ th group. While random effects may look like model coefficients, we are only interested in estimating their variances.

## Linear mixed effect model (LME) – data types:

- **Longitudinal data:** repeated measurements are performed on each individual unit. Several units are sampled. Number of observations may differ across units.
  - $y_{ti}$  - observation at time  $t$  for  $i$ -th individual.
  - $y_{ij}$  -  $i$ th observation of  $j$ th individual (if time aspect not relevant).
- **Hierarchical data structures:** data with two or more groups/levels of observations. Number of observations may differ across units.
  - $y_{ij}$  - observation for  $i$ -th company within  $j$ -th region.
  - $y_{ij}$  - observation for  $i$ -th student within  $j$ -th class.
- **Combined:** We can group observations at three levels (or more):
  - $y_{tij}$  - measurement at time period  $t$ , admin. region  $i$  within state  $j$ .
- Note how indices are ordered (left to right) from individual to highest level of aggregation. (alternative orderings exist in literature).

# Longitudinal data

- $N$  individual CS units are followed over time.
- The observation set  $\{y_{ti}, \mathbf{x}_{ti}\}$  denotes some  $i$ th individual observed at a given time period  $t$ . The number of observations in time may differ among CS units and observations may occur at different time points.

**Example:** For a medical study, we measure child's weight (plus other data) at birth and repeatedly over a period of one year. For some  $y_{ti}$  observation, index  $t$  denotes days from birth. Due to doctor visit scheduling, children are weighted at different  $t$  "values". Typically, the number of doctor visits (observations) differs across children. Children in the study are born on different dates (say, Jan 2015 - Oct 2019).

Example extends easily to economic environment (we can follow newly founded companies, etc.).

# Hierarchical data structures

- Nested/hierarchical structure of the LME model:
  - Individual units  $i$  (Level 1) are nested
  - within  $j$  groups (Level 2) with group-specific observation sizes  $n_j$ .
- One or more coefficient(s) can vary across groups (“random effects”).

## **LME: Longitudinal vs. hierarchical data structures:**

- Essentially, the same nesting/hierarchical framework applies to longitudinal data and their LME-based analysis:
  - Observations at time  $t$  (Level 1) are nested
  - within  $j$  individual units (Level 2).
  - If appropriate, individual units can be nested in groups (Level 3) ...



# Introduction

- Mixed models are called “mixed”, because the  $\beta$ -coefficients are a mix of fixed parameters and random variables
- Terms “fixed” and “random” have specific meaning for LMEs:
  - A fixed coefficient is an unknown constant to be estimated.
  - A random coefficient varies from “group” to “group”.  
By “group”, we mean Level 2 aggregation, if data have 2 levels.
    - coefficients vary among schools (Level 2), not within school.
    - coeffs. vary across individuals (Level 2), not over time (Level 1).
- LME models can have some added complexity:
  - Multiple levels of nesting
  - Crossed random effects
  - Correlations between different random coefficients.
- Random coefficients are not estimated, but they can be predicted.

# LME model example

- Data:  
London Education Authority Junior School Project dataset,
  - we have 887 students ( $i$ ) in 48 different schools ( $j$ ),
  - we want to predict 5th-year math scores.
- We may start by ignoring the school grouping and any possible regressors – we have a trivial model (*single-mean* model):

$$\text{math5}_{ij} = \beta_0 + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where  $M = 48$  and  $n_j$  differ among schools,  $\text{math5}_{ij}$  is the observed math score of  $i$ -th student at school  $j$ ,  $\beta_0$  is the mean math score across our population (being sampled) and  $\varepsilon_{ij}$  is the individual deviation from overall mean.

Population mean math score & the variance of  $\varepsilon$  are estimated by taking their sample counterparts. Any “school effect” is ignored.

- The school effect (differences among schools) may be incorporated in the model by allowing the mean of each school to be represented by a separate parameter (*fixed effect*)

$$\text{math5}_{ij} = \beta_{0j} + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, M, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where  $\beta_{0j}$  is the school-specific mean math score and  $\varepsilon_{ij}$  is the individual deviation from the school-specific mean.

- R syntax: `lm(math5 ~ School-1, data=...)`  
 $\Rightarrow M = 48$  school-specific intercepts are estimated.
- Using the terminology of LME,  $\beta_{0j}$  are fixed. Hence:
  - **Estimated intercepts only model (refer to) the specific sample** of schools, while -usually- the main interest is in the population from which the sample was drawn.
  - OLS regression does not provide an estimate of the between-school variability, which is also of central interest.

- *Random effects* approach: LME model can solve the above problem by treating school effects as random variations around population mean.
- Ordinary model (with *fixed effects*) can be reparametrized as:

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$y_{ij} = \beta_0 + (\beta_{0j} - \beta_0) + \varepsilon_{ij},$$

*Random effect:*  $u_{0j} = \beta_{0j} - \beta_0$  is the school-specific deviation from overall mean  $\beta_0$ . It can be used to replace the *fixed effect*  $\beta_{0j}$ :

$$u_{0j} = \beta_{0j} - \beta_0 \quad \Rightarrow \quad \beta_{0j} = \beta_0 + u_{0j}. \text{ Hence:}$$

$$y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}.$$

- $u_{0j}$  is a random variable, specific for the  $j$ -th school, with zero mean and unknown variance  $\sigma_u^2$ .

$u_{0j}$  is a *random effect*, associated with the particular sample units (schools are selected at random from the population).

- LME model with *random effects* (on the intercept) is given as:

$$y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}, \quad u_{0j} \sim N(0, \sigma_u^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2),$$

and we assume  $u_{0j}$  are *iid* and independent from  $\varepsilon_{ij}$ .

- Observations within the same school share the same random effect  $u_{0j}$ , hence are positively “correlated” with  $\text{ICC} = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2)$  (see ICC discussion on a separate slide).
- This *random effects* model has three parameters:  $\beta_0$ ,  $\sigma_u^2$  and  $\sigma_\varepsilon^2$ . (regardless of  $M$ , the number of schools).
- Note that the *random effect*  $u_{0j}$  “looks like” a coefficient, but we are only interested in estimating  $\sigma_u^2$ .
- However, upon observed data (and estimated model), we do make predictions using fitted values of  $\hat{u}_j$ .

- Exogenous regressors can be used in LMEs (like in LRMs).  
For example, `math5` grades depend on `math3` (3<sup>rd</sup> year grades).

$$\text{math5}_{ij} = (\beta_0 + u_{0j}) + \beta_1 \text{math3}_{ij} + \varepsilon_{ij}$$

alternative notation:

$$\text{math5}_{ij} = \beta_0 + \beta_1 \text{math3}_{ij} + u_{0j} + \varepsilon_{ij}$$

alternative notation:

$$\text{Level 1: } \text{math5}_{ij} = \beta_{0j} + \beta_1 \text{math3}_{ij} + \varepsilon_{ij}$$

$$\text{Level 2: } \beta_{0j} = \beta_0 + u_{0j}$$

- Intercept has a random effect, given the  $u_{0j}$  element.
- Slope of the regression line for each school is fixed at  $\beta_1$ .  
...`math3` has a *fixed effect*.

- **ICC:** Intra class correlation in a LME regression model:

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

- Describes how strongly units in the same group are “correlated”.
- While interpreted as a type of correlation, ICC operates on groups, rather than paired observations.
- See [link](#) for relation between ICC and actual correlation.

- **ICC:** Intra class correlation in a LME regression model:

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}$$

- Example:  $\text{math5}_{ij} = \beta_0 + \beta_1 \text{math3}_{ij} + u_{0j} + \varepsilon_{ij}$ ,  
where  $\sigma_u^2 = \text{var}(u_{0j})$  and  $\sigma_\varepsilon^2 = \text{var}(\varepsilon_{ij})$ .
- ICC: “correlation” between **math5** observations (randomly chosen) within a given school.
- ICC has another useful interpretation: Say,  $\text{ICC} = 0.6$  in our **math5**<sub>ij</sub> example. Hence, differences between schools explain 60% of “remaining” variance – i.e. after the variance explained by fixed effects (i.e. by **math3**<sub>ij</sub>) is subtracted.



# LME model: random effects on intercept and slope

- If teaching is different from school to school, it would make sense to have different slopes for each of the schools.

Instead of using *fixed effects* on slopes (interaction terms `math3:School`), we use random slopes:  $u_{1j} = \beta_{1j} - \beta_1$ .

$$\text{math5}_{ij} = (\beta_0 + u_{0j}) + (\beta_1 \text{math3}_{ij} + u_{1j} \text{math3}_{ij}) + \varepsilon_{ij},$$

alternative notation:

$$\text{math5}_{ij} = \underbrace{\beta_0 + \beta_1 \text{math3}_{ij}}_{\text{fixed}} + \underbrace{u_{0j} + u_{1j} \text{math3}_{ij}}_{\text{random}} + \varepsilon_{ij},$$

- We can test whether this extra complexity is justified.
- $u_{0j}$  and  $u_{1j}$  are often correlated, their independence can be tested.
- Fitted values of  $\text{math5}_{ij}$  can be produced, along with  $\hat{u}_{0j}$  and  $\hat{u}_{1j}$ .

- LME model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad \mathbf{u} \sim N(\mathbf{0}, \mathbf{G}) \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R}),$$

where:

$\mathbf{X}$  is a  $(n \times k)$  matrix,  $k$  is the number of *fixed effects*,  
 $\mathbf{Z}$  is a  $(n \times p)$  matrix,  $p$  is the number of *random effects*,  
 $\mathbf{G}$  is a  $(p \times p)$  variance-covariance matrix of the *random effects*,  
 $\mathbf{R}$  is a  $(n \times n)$  variance-covariance matrix of errors.

- Independence between  $\mathbf{u}$  and  $\boldsymbol{\varepsilon}$  is assumed,
  - Often,  $\mathbf{R} = \sigma_{\varepsilon}^2 \mathbf{I}_n$  is assumed,
  - $\mathbf{G}$  is diagonal if *random effects* are mutually independent.
- Estimation: MLE, RMLE, penalized least squares  
<https://www.jstatsoft.org/article/view/v067i01/0>

For a single random effect (on the intercept):

- `{nlme}` package:

```
lme( y ~ x + z, random = ~ 1 | g, data = df )
```

- `{lme4}` package:

```
lmer( y ~ x + z + ( 1 | g ), data = df )
```

- where `y` is the response variable with predictors `x` and `z`, and grouping factor variable `g`.

For random effects on the intercept and x:

- `{nlme}` package:

```
lme( y ~ x + z, random = ~ 1 + x | g, data = df )  
lme( y ~ x + z, random =      ~ x | g, data = df )
```

- `{lme4}` package:

```
lmer( y ~ x + z + ( 1 + x | g ), data = df )  
lmer( y ~ x + z +      ( x | g ), data = df )
```

For uncorrelated random effects on the intercept and  $x$ :

- `{nlme}` package:

```
lme( y ~ x + z, random = list ( g = pdDiag ( ~ x ) ) ,  
    data = df )
```

- `{lme4}` package:

```
lmer( y ~ x + z + ( 1 | g ) + ( 0 + x | g ) , data = df )  
lmer( y ~ x + z + ( x || g ) , data = df )
```

Different types of LME models exist:

- LME models with (multilevel) nested effects,
- LME models with crossed effects,
- Complex behavior of the error term in LME models can be addressed (heteroscedasticity and serial correlation).
- LME models with non-Gaussian dependent variables (binary, Poisson, etc.).

**Multi-level model example:** For 17 years, we follow a total of 86 individual states organized within 9 “global-level” regions (e.g. South America, Europe, Middle East, etc.).

- $\text{GDP}_{tij}$  represents individual GDP per capita measurements for:  
 $t$ -th time period, e.g. with values ( $t = 2000, \dots, 2016$ ).  
 $i$ -th state nested within region  $j$  ( $i = 1, \dots, M_j$ ),  
 $j$ -th region ( $j = 1, \dots, 9$ ),
- We fit  $\text{GDP}$  as a function of productivity  $\text{P}$  and unemployment  $\text{U}$ .  
States are nested in regions, we have 2 levels of random intercepts:  
 $u_{0i(j)}$  for each state (within a region),  
 $v_{0j}$  for the regions,  
random slopes can be added as well.
- $\text{GDP}_{tij} = \beta_0 + \beta_1 \text{P}_{tij} + \beta_2 \text{U}_{tij} + u_{0i(j)} + v_{0j} + \varepsilon_{tij}$ .

# LME: R syntax

For intercept varying among g1 and g2 within g1:

For intercept & x varying among g1 and g2 within g1:

- {nlme} package:

```
lme( y ~ x + z, random = ~ 1 | g 1 / g 2 , data = df )  
lme( y ~ x + z, random = ~ x | g 1 / g 2 , data = df )
```

- {lme4} package:

```
lmer( y ~ x + z +( 1 | g 1 / g 2 ), data = df )  
lmer( y ~ x + z +( x | g 1 / g 2 ), data = df )
```



Crossed *random effects* example:

- Grunfeld (1958) analyzed data on 10 large U.S. corporations, collected annually from 1935 to 1954 to investigate how investment  $I$  depends on market value  $M$  and capital stock  $C$ .
- Here, we want *random effects* for a given firm and year. We want the year effect to be the same across all firms, i.e. not nested within firms.
- $I_{ti} = \beta_0 + \beta_1 M_{ti} + \beta_2 C_{ti} + u_{0i} + v_{0t} + \varepsilon_{ti}$ .  
where  $i = 1, \dots, 10$  and  
firms are followed over  $t = 1, \dots, 20$  years.  
(the usual “ $it$ ” index ordering can be used as well)

For intercept varying among g1 and g2

- {lme4} package:

```
lmer( y ~ x + z + ( 1 | g 1 ) + ( 1 | g 2 ), data = df )
```

# LME with heteroscedasticity and serial correlation

`{nlme}` package:

- Heteroscedastic residual variance at level 1:

```
lme( y ~ x + z, random = ~ 1 | g ,  
     weights = varIdent ( form = ~ 1 | g ) , data )
```

- Autoregressive `ar(1)` residuals:

```
lme( y ~ x + z, random = ~ 1 | g ,  
     correlation = corAR1 ( form = ~ time ) , data )
```

- General residuals:

```
lme( y ~ x + z, random = ~ 1 | g ,  
     weights = varIdent ( form = ~ 1 | g ) ,  
     correlation = corAR1 ( form = ~ time ) , data )
```

# LME models – references, R packages

- {lme4} package  
<https://www.jstatsoft.org/article/view/v067i01/0>
- {nlme} package  
<https://cran.r-project.org/web/packages/nlme/nlme.pdf>
- <https://www.r-bloggers.com/2017/12/linear-mixed-effect-models-in-r/>
- <https://rpsychologist.com/r-guide-longitudinal-lme-lmer>
- Finch, Bolin, Kelley: Multilevel Modeling Using R (2014).